

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.**

Application Serial Number: 10/S72,638  
Source: IFWP  
Date Processed by STIC: 02/10/2007

***ENTERED***



IFWP

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/10/572,638**

**DATE: 02/10/2007**  
**TIME: 09:56:30**

**Input Set : N:\DA\pte.dat.txt**  
**Output Set: N:\CRF4\02102007\J572638.raw**

3 <110> APPLICANT: DUKE UNIVERSITY  
 5 <120> TITLE OF INVENTION: CONSENSUS/ANCESTRAL IMMUNOGENS  
 7 <130> FILE REFERENCE: 01579-1093  
 9 <140> CURRENT APPLICATION NUMBER: 10/572,638  
 C--> 10 <141> CURRENT FILING DATE: 2006-03-17  
 12 <150> PRIOR APPLICATION NUMBER: PCT/US2004/03097  
 13 <151> PRIOR FILING DATE: 2004-09-17  
 15 <150> PRIOR APPLICATION NUMBER: 60/604,722  
 16 <151> PRIOR FILING DATE: 2004-08-27  
 18 <150> PRIOR APPLICATION NUMBER: 60/503,460  
 19 <151> PRIOR FILING DATE: 2003-09-17  
 21 <160> NUMBER OF SEQ ID NOS: 321  
 23 <170> SOFTWARE: PatentIn Ver. 3.3  
 25 <210> SEQ ID NO: 1  
 26 <211> LENGTH: 866  
 27 <212> TYPE: PRT  
 28 <213> ORGANISM: Human immunodeficiency virus  
 30 <400> SEQUENCE: 1  
 31 Met Arg Val Met Gly Ile Gln Arg Asn Cys Gln His Leu Trp Arg Trp  
 32 1 5 10 15  
 34 Gly Thr Met Ile Leu Gly Met Leu Met Ile Cys Ser Ala Ala Glu Asn  
 35 20 25 30  
 37 Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Asn  
 38 35 40 45  
 40 Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val  
 41 50 55 60  
 43 His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro  
 44 65 70 75 80  
 46 Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys  
 47 85 90 95  
 49 Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp  
 50 100 105 110  
 52 Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu  
 53 115 120 125  
 55 Asn Cys Thr Asn Val Arg Asn Val Ser Ser Asn Gly Thr Glu Thr Asp  
 56 130 135 140  
 58 Asn Glu Glu Ile Lys Asn Cys Ser Phe Asn Ile Thr Thr Glu Leu Arg  
 59 145 150 155 160  
 61 Asp Lys Lys Gln Lys Val Tyr Ala Leu Phe Tyr Arg Leu Asp Val Val  
 62 165 170 175  
 64 Pro Ile Asp Asp Lys Asn Ser Ser Glu Ile Ser Gly Lys Asn Ser Ser  
 65 180 185 190  
 67 Glu Tyr Tyr Arg Leu Ile Asn Cys Asn Thr Ser Ala Ile Thr Gln Ala

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/572,638

DATE: 02/10/2007

TIME: 09:56:30

Input Set : N:\DA\pto.da.txt

Output Set: N:\CRF4\02102007\J572638.raw

68	195	200	205
70	Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro		
71	210	215	220
73	Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly Thr		
74	225	230	235
76	Gly Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys		240
77	245	250	255
79	Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu		
80	260	265	270
82	Glu Ile Ile Ile Arg Ser Glu Asn Ile Thr Asn Asn Ala Lys Thr Ile		
83	275	280	285
85	Ile Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn		
86	290	295	300
88	Asn Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Gln Ala Phe Tyr		
89	305	310	315
91	Ala Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile		320
92	325	330	335
94	Ser Arg Thr Lys Trp Asn Lys Thr Leu Gln Gln Val Ala Lys Lys Leu		
95	340	345	350
97	Arg Glu His Phe Asn Asn Lys Thr Ile Ile Phe Lys Pro Ser Ser Gly		
98	355	360	365
100	Gly Asp Leu Glu Ile Thr Thr His Ser Phe Asn Cys Gly Gly Glu Phe		
101	370	375	380
103	Phe Tyr Cys Asn Thr Ser Gly Leu Phe Asn Ser Thr Trp Met Phe Asn		
104	385	390	395
106	Gly Thr Tyr Met Phe Asn Gly Thr Lys Asp Asn Ser Glu Thr Ile Thr		400
107	405	410	415
109	Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Gly Val Gly		
110	420	425	430
112	Gln Ala Met Tyr Ala Pro Pro Ile Glu Gly Lys Ile Thr Cys Lys Ser		
113	435	440	445
115	Asn Ile Thr Gly Leu Leu Leu Ile Thr Arg Asp Gly Gly Asn Asn Ser		
116	450	455	460
118	Asn Lys Asn Lys Thr Glu Thr Phe Arg Pro Gly Gly Asp Met Arg		
119	465	470	475
121	Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu		480
122	485	490	495
124	Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Glu Arg		
125	500	505	510
127	Glu Lys Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly		
128	515	520	525
130	Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln		
131	530	535	540
133	Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Ser Asn Leu Leu		
134	545	550	555
136	Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly		560
137	565	570	575
139	Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu Lys		
140	580	585	590

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/572,638

DATE: 02/10/2007

TIME: 09:56:30

Input Set : N:\DA\pto.da.txt

Output Set: N:\CRF4\02102007\J572638.raw

142 Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys  
 143 595 600 605  
 145 Thr Thr Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Asp  
 146 610 615 620  
 148 Glu Ile Trp Asp Asn Met Thr Trp Met Glu Trp Glu Arg Glu Ile Ser  
 149 625 630 635 640  
 151 Asn Tyr Thr Asp Ile Ile Tyr Arg Leu Ile Glu Glu Ser Gln Asn Gln  
 152 645 650 655  
 154 Gln Glu Lys Asn Glu Gln Glu Leu Leu Ala Leu Asp Lys Ala Ser Leu  
 155 660 665 670  
 157 Trp Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe  
 158 675 680 685  
 160 Ile Met Ile Val Gly Gly Leu Ile Gly Leu Arg Ile Val Phe Ala Val  
 161 690 695 700  
 163 Leu Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe  
 164 705 710 715 720  
 166 Gln Thr Leu Ile Pro Asn Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile  
 167 725 730 735  
 169 Glu Glu Glu Gly Glu Gln Gly Arg Asp Arg Ser Ile Arg Leu Val  
 170 740 745 750  
 172 Asn Gly Phe Leu Ala Leu Ala Trp Asp Asp Leu Arg Ser Leu Cys Leu  
 173 755 760 765  
 175 Phe Ser Tyr His Arg Leu Arg Asp Phe Ile Leu Ile Ala Ala Arg Thr  
 176 770 775 780  
 178 Val Glu Leu Leu Gly Arg Arg Ser Leu Arg Gly Leu Gln Lys Gly Trp  
 179 785 790 795 800  
 181 Glu Ala Leu Lys Tyr Leu Gly Asn Leu Leu Gln Tyr Trp Gly Gln Glu  
 182 805 810 815  
 184 Leu Lys Asn Ser Ala Ile Ser Leu Leu Asp Thr Thr Ala Ile Ala Val  
 185 820 825 830  
 187 Ala Glu Gly Thr Asp Arg Val Ile Glu Ile Val Gln Arg Ala Cys Arg  
 188 835 840 845  
 190 Ala Ile Leu Asn Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu Arg Ala  
 191 850 855 860  
 193 Leu Leu  
 194 865  
 197 <210> SEQ ID NO: 2  
 198 <211> LENGTH: 2607  
 199 <212> TYPE: DNA  
 200 <213> ORGANISM: Human immunodeficiency virus  
 202 <400> SEQUENCE: 2  
 203 gccaccatgc gcgtgatggg catccagcgc aactgccagc acctgtggcg ctggggcacc 60  
 204 atgatcctgg gcatgctgat gatctgctcc gcccggaga acctgtgggt gaccgtgtac 120  
 205 tacggcgtgc ccgtgtggaa ggaggccaac accaccctgt tctgcgcctc cgacgccaag 180  
 206 gcctacgaca ccgaggtgca caacgtgtgg gccacccacg cctgcgtgcc caccgacccc 240  
 207 aaccccccagg agatcgtgct ggagaacgtg accgagaact tcaacatgtg gaagaacaac 300  
 208 atggtgagac agatgcacga ggacatcatc tccctgtggg accagtccct gaagccctgc 360  
 209 gtgaagctga cccccctgtg cgtgaccctg aactgcacca acgtgcgcaa cgtgtccctcc 420  
 210 aacggcacccg agaccgacaa cgaggagatc aagaactgct cttcaacat caccaccgag 480

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/572,638

DATE: 02/10/2007

TIME: 09:56:30

Input Set : N:\DA\pto.da.txt

Output Set: N:\CRF4\02102007\J572638.raw

211 ctgcgcgaca agaagcagaa ggtgtacgcc ctgttctacc gcctggacgt ggtccccatc 540  
 212 gacgacaaga actccctcgatctccggc aagaactcct ccgagtacta ccgcctgatc 600  
 213 aactgcaaca cctccggccat cacccaggcc tgccccaagg tgccttcga gccatcccc 660  
 214 atccactact gcccggccgc cggcttcgcc atccctgaagt gcaacgacaa gaagttcaac 720  
 215 ggcaccggcc cctgcaagaa cgtgtccacc gtgcagtgc cccacggcat caagcccgtg 780  
 216 gtgtccaccc agctgctgct gaacggctcc ctggccgagg aggagatcat catccgtcc 840  
 217 gagaacatca ccaagaarccgaaagaccatc atcgtgcagc tgaacgagtc cgtggagatc 900  
 218 aactgcaccc gccccaaacaa caaacaccgc aagtccatcc acatcgcccc cggccaggcc 960  
 219 ttctacgcca cccggagat catcgccgac atccggccagg cccactgcaa catctccgc 1020  
 220 accaagtggc acaagaccct gcagcagggtg gccaagaagc tgcgcgagca cttaacaac 1080  
 221 aagaccatca tcttaaagcc ctccctccggc ggcgacctgg agatcacccac ccactccttc 1140  
 222 aactgcggcg gcgagttctt ctactgcaac acctccggcc tgcgcgagatc cacttggatg 1200  
 223 ttcaacggca cctacatgtt caacggccacc aaggacaact cccggccat caccctgccc 1260  
 224 tgccgcatac agcagatcat caacatgtgg caggcgtgg gccaggccat gtacgcccc 1320  
 225 cccatcgagg gcaagatcac tgcgcgatcc aacatcaccgc gcctgctgt gacccggccac 1380  
 226 ggcggcaaca actccaaacaa gaacaagacc gagaccttcc gccccggccgg cggcgcacatg 1440  
 227 cgcgacaact ggcgcctccgatctgtacaagg tacaagggtgg tgaagatcga gcccctggc 1500  
 228 gtggcccccaccaaggccaa ggcgcgcgtg gtggagcgcg agaagcgcgc cgtgggcattc 1560  
 229 ggcgcgtgtt tccctgggtt ctggggccgc gccgggttcca cccatggggccgc cgcctccatc 1620  
 230 accctgaccc tgcaggccccgc cccatgggtgt tccggcatcg tgcagcagca gtccaaacctg 1680  
 231 ctgcgcgcca tcgagggccca gcagcacccgc ctgcagctga cccgtgtggg catcaagcag 1740  
 232 ctgcaggccccgc cccatgggtgtt ccgtggagccgc tacatgttacc accagcagatc gctgggcattc 1800  
 233 tggggctgtt ccggcaagct gatctgcacc accaacgtgc cctggaaactc ctctgggtcc 1860  
 234 aacaagtccc aggacgagat ctgggacaac atgacacttgc tggagtggga ggcgcgagatc 1920  
 235 tccaaactaca ccgcacatcat ctaccgcctt atcgaggatg cccagaacca gcaggagaag 1980  
 236 aacgaggcagg agctgctggc ctggacaag tggggctccc tgcgcgatc gttcgacatc 2040  
 237 accaaactggc tgcgttacat caagatcttc atcatgtatcg tggggccctt gatcgccctg 2100  
 238 cgcacatgtgt tgcgcgtgtt ccgtggatccgc aaccggcgtgc gccaggccat ctcctccctg 2160  
 239 tccttccaga ccctgtatccc caaccccccgc ggcgcgcacc gccccggagg catcgaggag 2220  
 240 gagggcggccg agcaggccgcg cggccgcctt atccgcctgg tgaacgggtt cctggccctg 2280  
 241 gcctgggacg acctgcgcctc cctgtgcctt ttcccttacc accgcctgcg cgacttcatc 2340  
 242 ctgatcgccg cccgcaccgt ggagctgtgc ggcgcgcgtt ccctgcgcgg cctgcagaag 2400  
 243 ggctgggagg ccctgaagta cctgggcaac ctgcgtgcgtt actggggcca ggagctgaag 2460  
 244 aactccgcaca tctccctgtt ggacaccacc gccatgcgcg tggccgaggcc caccgaccgc 2520  
 245 gtgatcgaga tcgtgcagcg cgcctgcgcg gccatccttga acatccccccgc cgcgcattccgc 2580  
 246 caggcccttgg agcgcgcctt gctgtaa 2607  
 249 <210> SEQ ID NO: 3  
 250 <211> LENGTH: 2535  
 251 <212> TYPE: DNA  
 252 <213> ORGANISM: Human immunodeficiency virus  
 254 <400> SEQUENCE: 3  
 255 ggcgcctatgc gcgtgtatggg catccgcgc aactgcgcgc agtgtgtggat ctggggccatc 60  
 256 ctggccatgc ggatgtat gatctgcgttcc gtgggtggca acctgtgggtt gaccgtgtac 120  
 257 tacggcgtgc ccgtgtggaa ggaggccaa accaccctgt tctgcgcctc cgacgcacaa 180  
 258 gcctacgagc gcgagggtgc caacgtgtgg gcccacccacgc cctgcgtgc caccgcaccc 240  
 259 aaccccccagg agatgggtgtt ggagaaacgtg accgagaact tcaacatgtt gaagaacgac 300  
 260 atgggtggacc agatgcacgc ggcacatcatc tccctgtgg accagtccctt gaagccctgc 360  
 261 gtgaagctga ccccccgtt cgtgcaccgc gccatccttga acatccccccgc cgcgcattccgc 420  
 262 aacacccatca acggcgagat gaagaactgc tccttcaaca tcaccaccgc gctgcgcac 480

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/10/572,638**

**DATE: 02/10/2007**  
**TIME: 09:56:30**

**Input Set : N:\DA\pto.da.txt**  
**Output Set: N:\CRF4\02102007\J572638.raw**

263 aagaagaaga aggagtacgc cctgttctac cgccctggaca tcgtccccct gaacgagaac 540  
 264 tcctcccgagt accgcctgat caactgcaac acctccgcca tcacccaggc ctgccccaaag 600  
 265 gtgtccttcg accccatccc catccactac tgccggcccg ccggctacgc catcctgaag 660  
 266 tgcaacaaca agaccccaa cggcacccgc ccctgcaaca acgtgtccac cgtgcagtgc 720  
 267 acccacggca tcaagcccggt ggtgtccacc cagctgctgc tgaacggctc cctggccgag 780  
 268 gaggagatca tcatccgctc cgagaacctg accgacaacg ccaagaccat catcggtcag 840  
 269 ctgaaacggagt ccgtggagat cgtgtgcacc cgccccaaca acaacacccg, caagtccatg 900  
 270 cycatcgcc cccggccagac cttctacgac accggcgaca tcatcgccgca catccggccag 960  
 271 gcccactgca acatctccga ggacaagtgg aacaagaccc tgcagcagggt ggccgagaag 1020  
 272 ctgggcaaggc acttcccaa caagaccatc accttcgagc cctccctccgg cggcgacctg 1080  
 273 gagatcacca cccactcctt caactgccc ggcgagttct tctactgcaaa cacctccaag 1140  
 274 ctgttcaact ccacccatcaa caacaacacc aactccaact ccaccatcac cctggccctgc 1200  
 275 cgcataaggc agatcatcaa catgtggcag ggcgtggcc aggccatgta cgcccccccc 1260  
 276 atcgccggca acatcacctg caagtccaaatcaccggcc tgctgctgac cgcgcacggc 1320  
 277 ggcaaggaga acaccacccgac gacctccgc cccggccggcg ggcacatgcg cgacaactgg 1380  
 278 cgctccgagc tgtacaagtgta caagggtggg gagatcaagc ccctggccgt ggccccccacc 1440  
 279 gaggccaaaggc gccgcgtggt ggagcgcgcg aagcgcgcgc tgggcctggg cgccgtgttc 1500  
 280 ctgggcttcc tgggcgcgcg cggctccaccatggcgcgcg cctccatcac cctgaccgtg 1560  
 281 .caagccgcg agctgtgtc cggcatcgatc cagcagcagg tcaacactgtc ggcgcgcac 1620  
 282 gaggcccgac agcacatgtc gcagctgacc gtgtggggca tcaaggcagct gcaggccgc 1680  
 283 gtgctggcca tggagcgccta cctgaaggac cagcagctgc tggcatctg gggctgtcc 1740  
 284 ggcaagctga tctgcaccac cgcgtgccc tggaaactcct cctgtccaa caagtccctg 1800  
 285 gacgacatct gggacaacat gacctggatg gagtggacc ggcagatctc caactacacc 1860  
 286 gacaccatct accgcctgct ggaggagtcc cagaaccagc aggagaagaa cgagcaggac 1920  
 287 ctgctggccc tggactcctg ggagaacctg tggaaactgtt tgcacatcac caactggctg 1980  
 288 tggtacatca agatottcat catgatcgat ggcggcctga tggcctgcg catcatctc 2040  
 289 gccgtgtgtt ccatcgtaa cgcgtgcgc cagggtact cccctgtc ttccagacc 2100  
 290 ctgaccccca acccccgccg ccccgacccgc ctggagcgc tggaggagga gggccggcag 2160  
 291 caggaccgcg accgcctccat cgcctgggt tccggcttcc tggccctggc ctggacgac 2220  
 292 ctgcgtccccc tggcgtgtt ctcttaccac cgcctgcgcg acttcattct gatcgccgc 2280  
 293 cgcaccgtgg agctgtgtgg cgcgtctcc ctgcgcggcc tgcagcgcgg ctgggaggcc 2340  
 294 ctgaagtacc tgggcctccct ggtgcagtac tggggccagg agctgaagaa gtccgcac 2400  
 295 tccctgtgg acaccatcgc catgcgcgtg gccgaggccgac cgcaccgcacatcgaggtg 2460  
 296 gtgcagcgcg cctgcgcgc catcctgaac atccccggcc gcatccgcac gggcttcgag 2520  
 297 gccgcctgc tgtaa 2535  
 300 <210> SEQ ID NO: 4  
 301 <211> LENGTH: 2535  
 302 <212> TYPE: DNA  
 303 <213> ORGANISM: Human immunodeficiency virus  
 305 <400> SEQUENCE: 4  
 306 gcccgcacatgc gcgtgtgtgg catcctgcgc aactgccagc agtgtggat ctggggcato 60  
 307 ctgggcttct ggatgtgtat gatctgcaac gtgtgggc acctgtgggt gaccgtgtac 120  
 308 tacggcgtgc ccgtgtggaa ggaggccaa accaccctgt tctgcgcctc cgacgcac 180  
 309 gcctacgaga aggagggtgc caacgtgtgg gccacccacg cctgcgtgccc caccgcac 240  
 310 aaccccccagg agatgtgtc ggagaacgtg accgagaact tcaacatgtg gaagaacgac 300  
 311 atgggtggacc agatgcacga ggacatcatc tccctgtggg accagtccct gaagccctgc 360  
 312 gtgaagctga ccccccgtgc cgtgaccctg aactgcccgc acgtgaccaa cgccaccaac 420  
 313 aacacccatca acgaggagat caagaactgc tccttcaaca tcaccaccgac gctgcgcac 480  
 314 aagaagaaga aggtgtacgc cctgttctac cgccctggaca tcgtccccct gaacgagaac 540

VERIFICATION SUMMARY  
PATENT APPLICATION: US/10/572,638

DATE: 02/10/2007  
TIME: 09:56:31

Input Set : N:\DA\pto.da.txt  
Output Set: N:\CRF4\02102007\J572638.raw

L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:2428 M:283 W: Missing Blank Line separator, <400> field identifier  
L:2429 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (27) SEQUENCE:  
L:14185 M:283 W: Missing Blank Line separator, <400> field identifier  
L:14186 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (149) SEQUENCE: